CURRENT MICROBIOLOGY Vol. 48 (2004), pp. 77–81
DOI: 10.1007/s00284-003-4132-3

Current Microbiology
An International Journal

Sequence Analysis of DNA Fragments from the Genome of the Primary Endosymbiont of the Whitefly *Bemisia tabaci*

Linda Baumann,[1] MyLo Ly Thao,[1] C. Joel Funk,[2] Bryce W. Falk,[3] James C.K. Ng,[3] Paul Baumann[1]

[1]Microbiology Section, University of California, One Shields Avenue, Davis, CA, 95616-8665
[2]U.S. Department of Agriculture, Agricultural Research Service, Western Cotton Research Laboratory, Phoenix, AZ 85040
[3]Department of Plant Pathology, University of California, Davis, CA, 95616-8680

Received: 18 April 2003 / Accepted: 30 April 2003

Abstract. The whitefly *Bemisia tabaci* contains a primary prokaryotic endosymbiont housed within specialized cells in the body cavity. Two DNA fragments from the endosymbiont, totaling 33.3 kilobases, were cloned and sequenced. In total, 37 genes were detected and included the ribosomal RNA operon and genes for ribosomal RNA proteins. The guanine plus cytosine of the DNA was 30.2 mol%, different from that of endosymbionts of other plant sap-sucking insects.

Whiteflies are members of the suborder Sternorrhyncha that also includes aphids, psyllids, and mealybugs [8]. These insects share a number of properties in that they utilize plant sap as a diet and contain prokaryotic endosymbionts within specialized cells known as bacteriocytes. These cells form a structure, the bacteriome, which is within the body cavity of the insect [3]. The association between these primary (P−) endosymbionts and the insect appears to be the result of a single infection of an ancestral insect with different bacteria that are maternally transmitted to progeny [3, 9, 20]. The consequence of this vertical transmission is cospeciation between the insect and the endosymbiont. Studies performed primarily with aphids indicate that one of the functions of the endosymbiont is an upgrading of the insect diet by the synthesis of essential amino acids for the insect host [7, 9]. The P-endosymbionts of aphids, psyllids, and mealybugs have been designated as *Buchnera*, *Carsonella*, and *Tremblaya*, respectively [2, 16, 17]. The different origin of these organisms is also indicated by their closest known free-living relatives. *Buchnera* is related to *Escherichia coli* (γ-subdivision of the *Proteobacteria*), *Carsonella* to *Pseudomonas aeruginosa* (γ-subdivision), while *Tremblaya* is related to *Burkholderia mallei* (β-subdivision).

The genomes of *Buchnera* from three aphid species have been sequenced and found to have a size of 616–642 kb [12, 15, 19]. A striking property of these genomes is the conservation of gene order and the presence of genes encoding for enzymes for the synthesis of essential amino acids. DNA fragments of *Carsonella* (37 kb) and *Tremblaya* (64 kb) have also been sequenced and were found to have a number of properties differing from those of *Buchnera* [1, 5].

Recently we have initiated studies on the evolutionary relationships of whitefly endosymbionts. Our studies have extended past observations that have indicated a single origin of the whitefly P-endosymbionts [4, 13], and we have given them the designation *Candidatus* Portiera aleyrodidarum (designated here as *Portiera*). In the present study, we report on the genetic properties of a 33-kb DNA fragment of *Portiera* from the whitefly *Bemisia tabaci.* This insect may also contain other prokaryotes including a secondary endosymbiont, *Arsenophonus,* and *Chlamydia* [4, 6, 18, 21].

Materials and Methods

General methods. The methods used in these studies have been described in our past publications [1, 2, 5]. These include methods for total whitefly DNA purification, polymerase chain reaction (PCR), amplification of the 16S–23S rDNA fragment and its cloning into pBluescript (Stratagene, LaJolla, CA), restriction enzyme and Southern blot analysis, electroelution of DNA fragments from agarose gels and their cloning into λZAP (Stratagene). The nucleotide sequence of the DNA fragments was determined at the University of Arizona (Tucson) LSME sequencing facility or at the Arizona State University DNA Sequencing Laboratory (Tempe). For sequence determination of most of the DNA fragments, a double-stranded DNA nested deletion kit

Correspondence to: P. Baumann; *email:* pabaumann@ucdavis.edu

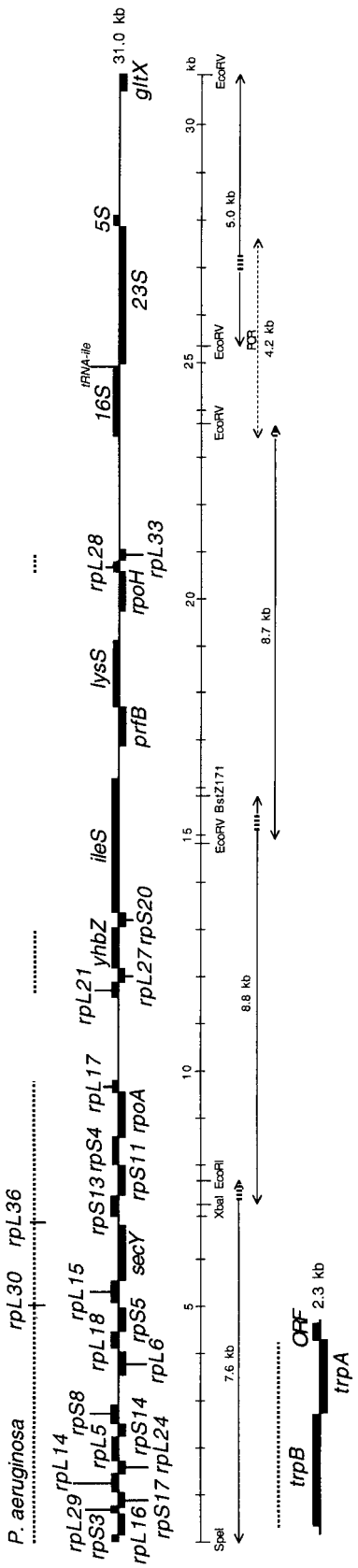

Fig. 1. Genetic maps of two sequenced DNA fragments of the P-endosymbiont of *Bemisia tabaci.* Thin solid line, DNA; thick solid line, coding regions; dashed line, order of genes in *P. aeruginosa;* double-headed arrows, cloned and sequenced overlapping DNA fragments; striped line on double-headed arrows, position of probe used to detect recombinants; double-headed arrow on dashed line, fragment obtained by PCR. The direction of transcription is left to right.

(Pharmacia, Piscataway, NJ) was used. In addition to T3 and T7 primers, in some cases, custom-made oligonucleotide primers were also designed for sequencing.

31-kb DNA fragment. We initially obtained the 4.2-kb 16S–23S rDNA fragment and determined its sequence (Fig. 1). Probes to 16S and 23S rDNA were constructed and used in restriction enzyme and Southern blot analysis of whitefly DNA. With the 16S rDNA probe, a predominant 8.7-kb and a minor 5.5-kb *Eco*RV fragment were detected. These were electroeluted from agarose gels and cloned into λZAP. The 8.7-kb DNA fragment corresponded to the P-endosymbiont, while the 5.5-kb DNA fragment contained a partial 16S rDNA sequence related to that of the "*Encarsia bacterium*" (96% identity) [21] and will not be further considered. With similar methods, additional DNA fragments upstream and downstream of the 16S–23S rDNA genes were obtained (Fig. 1). In all cases, only single bands were observed in restriction enzyme and Southern blot analysis.

2.3-kb DNA fragment. An initial 0.73-kb region of *trpB* (Fig. 1) was PCR amplified by using the degenerate primers previously described [10]. After the PCR fragment was sequenced, flanking regions were cloned using Partial Inverse PCR [11] and sequenced.

Analysis of the DNA. We used GeneJockey II (Biosoft, Ferguson, MO) to identify open reading frames (ORFs) and Blast searches (NCBI, Bethesda, MD) for proteins with amino acid sequence similarity. Alignment of amino acids was performed with Gap (Genetics Computer Group, Madison, WI). In comparative studies, sequences of *Pseudomonas aeruginosa* AE004091 and *Zymobacter palmae* AF211871 were also included.

Source of insects. *Bemisia tabaci* (biotype A) were reared in the laboratory of B. W. Falk (31.0-kb DNA fragment). For the 2.3-kb fragment, another strain of this species was reared in the laboratory of Joel Funk.

Accession numbers. The GenBank accession numbers for the 31.0- and 2.3-kb DNA fragments are AY268081 and AY268082, respectively. The accession number for the 5.5-kb DNA fragment from the relative of the "*Encarsia bacterium*" is AY271301.

Results and Discussion

A genetic map of the DNA fragments of *Portiera* from the whitefly *Bemisia tabaci* is presented in Fig. 1. The two DNA fragments were 31,043 and 2,256 nucleotides (nt) in length (combined length 33,299 nt). The guanine plus cytosine (G+C) content of the DNA was 30.2 mol%. A total of 33 open reading frames (ORFs) was detected, which corresponded to proteins in the databases. A list of the genes together with their protein designations and the % amino acid identity to homologous *P. aeruginosa* proteins is presented in Table 1. The comparisons are made with *P. aeruginosa* since this is

Table 1. Genes found in *Portiera*, the P-endosymbiont of *Bemisia tabaci*

Gene	Protein	% Amino acid identity[a]
Regulatory functions		
rpoH	RNA polymerase σ^{32} subunit; heat-shock response	57.6
Ribosomal RNA		
rrf	5S rRNA	78.3[b]
rrl	23S rRNA	81.2[b]
rrs	16S rRNA	83.5[b]
Ribosomal proteins		
rpS3[c]	Ribosomal protein S3 (*rpsC*)	47.1
rpS4	Ribosomal protein S4 (*rpsD*)	62.7
rpS5	Ribosomal protein S5 (*rpsE*)	59.2
rpS8	Ribosomal protein S8 (*rpsH*)	53.5
rpS11	Ribosomal protein S11 (*rpsK*)	66.7
rpS13	Ribosomal protein S13 (*rpsM*)	62.1
rpS14	Ribosomal protein S14 (*rpsN*)	59.4
rpS17	Ribosomal protein S17 (*rpsQ*)	48.2
rpS20	Ribosomal protein S20 (*rpsT*)	45.6
rpL5	Ribosomal protein L5 (*rplE*)	58.5
rpL6	Ribosomal protein L6 (*rplF*)	48.3
rpL14	Ribosomal protein L14 (*rplH*)	81.2
rpL15	Ribosomal protein L15 (*rplO*)	45.8
rpL16	Ribosomal protein L16 (*rplP*)	67.9
rpL17	Ribosomal protein L17 (*rplQ*)	59.5
rpL18	Ribosomal protein L18 (*rplR*)	52.2
rpL21	Ribosomal protein L21 (*rplU*)	47.6
rpL24	Ribosomal protein L24 (*rplX*)	50.0
rpL27	Ribosomal protein L27 (*rpmA*)	64.7
rpL28	Ribosomal protein L28 (*rpmB*)	59.2
rpL29	Ribosomal protein L29 (*rpmC*)	47.5
rpL33	Ribosomal protein L33 (*rpmG*)	55.1
tRNAs		
tRNA^{Ile}	Isoleucine tRNA	84.9[b]
Aminoacyl-tRNA synthetases		
gltX[c]	Glytamyl tRNA synthetase, catalytic subunit	61.0
ileS	Isoleucyl tRNA synthetase	53.8
lysS	Lysyl tRNA synthetase	51.2
RNA synthesis		
rpoA	RNA polymerase, α-subunit	66.5
Protein secretion		
secY	Membrane protein, secretion	46.5
Tryptophan biosynthesis		
trpA	Tryptophan synthase, α-subunit	44.0
trpB	Tryptophan synthase, β-subunit	69.0
Miscellaneous		
prfB	Peptide chain release factor	54.3[d]
yhbZ	Putative GTP-binding protein	42.9
ORF	Ferridoxin, truncated at the C-terminus	64.5

[a] Unless otherwise stated, the comparison is to *P. aeruginosa* proteins.
[b] Compared with RNAs.
[c] Partial sequence.
[d] Compared with *E. coli* protein.

the closest relative of *Portiera* for which the complete genome sequence has been determined [14]. Most of the genes encode for ribosomal RNA proteins. The order of these genes is similar to that of *P. aeruginosa* (Fig. 1). The DNA segment of *Portiera* from *rpS3* to *rpL17* differs from *P. aeruginosa* in lacking *rpL30* and *rpL36*. The amino acid sequence identity of the homologous proteins ranges from 42.9% for the putative GTP-binding protein to 81.2% for ribosomal protein L16 (Table 1). The 31-kb DNA fragment also contained genes for tRNA

Table 2. Summary of genetic characteristics of the P-endosymbionts of plant sap-sucking insects[a]

Endosymbiont designation	*Portiera*	*Carsonella*	*Buchnera*	*Tremblaya*
Insect host	whiteflies	psyllids	aphids	mealybugs
Proteobacterial group	γ	γ	γ	β
Free-living relative[b]	Pa	Pa	Ec	Bm
Size of analyzed DNA (kb)	33.3	37.1	641[c]	64.4
Moles % G+C content of DNA	30.2	19.9	26.3	57.1
% coding	71.2	>99.9	87.9	81.6
% G+C of intergenic spaces	23.9	−	19.1	57.4
Recognizable −35 −10 region preceding rRNA genes	−	−	+	−
Inverted repeats following rRNA genes	−	−	+	−
3′ end of 16S rDNA contains Shine-Dalgarno complement	+	−	+	+
Order of rRNA genes 16S–23S–5S	+	+	−[d]	+
tRNA between 16S and 23S rDNA	+	−	−[d]	−
Number of rRNA operons	1	1	1	2
Translational coupling the norm[e]	−	+	−	−
Increase A+T content in poorly conserved genes	+	++	+	−

[a] Interpretations based on analysis of sequence data [1, 5, 12].
[b] Pa = *Pseudomonas aeruginosa;* Ec = *Escherichia coli;* Bm = *Burkholderia mallei.*
[c] Full genome of *Buchnera* [12].
[d] 16S and 23S–5S genes are separated.
[e] Suggested from large number of gene pairs in which the initiation codon and the stop codon overlap [5].

synthases, the α-subunit of RNA polymerase, SecY, and σ^{32} subunit of RNA polymerase involved in the heat-shock response. In addition, the 2.3-kb DNA fragment contained the genes encoding for tryptophan synthase. There were three intergenic regions of substantial size that did not contain open reading frames. These are between *rpl17* and *rpl21* (1.6 kb), *rpL33* and *16S* (2.4 kb), and *5S* and *gltX* (2.9 kb) (Fig. 1). The G+C content of all the combined intergenic regions is 23.9%, considerably lower that that of the coding regions (32.8 mol%). The nt sequence of *Portiera* 16S and 23S rDNA has an identity of 83.5% and 81.2%, respectively, to the genes from *P. aeruginosa,* and 87.0% and 82.9% identity to the genes from *Zymobacter palmae.* The latter is the closest known relative of *Portiera.*

Table 2 compares some of the genetic properties of *Portiera* with that of other endosymbionts of plant sap sucking insects. These comparisons are preliminary, and additional differences may be found since only the genome of *Buchnera* has been completely sequenced. *Portiera* differs from the other endosymbionts in its G+C content (30.2 mol%) and the presence of a tRNA (isoleucine) between the 16S and 23S rRNA genes. It also appears to have the largest proportion of DNA that does not code for proteins, a conclusion that needs additional sequence data for verification.

ACKNOWLEDGMENTS

This material is based on work supported by National Science Foundation Awards DEB-9978518 (N. A. Moran, P. Baumann) and the University of California Agriculture Experiment Station (P. Baumann).

Literature Cited

1. Baumann L, Thao ML, Hess JM, Johnson MW, Baumann P (2002) The genetic properties of the primary endosymbionts of mealybugs differ from those of other endosymbionts of plant sap-sucking insects. Appl Environ Microbiol 68:3198–3205
2. Baumann P, Moran MA, Baumann L (2000) Bacteriocyte-associated endosymbionts of insects. In: Dworkin M (ed) The prokaryotes. New York: Springer, http://link.springer.de/link/service/books/10125
3. Buchner P (1965) Endosymbiosis of animals and plant microorganisms. New York: Interscience, pp 207–432
4. Clark MA, Baumann L, Munson MA, Baumann P, Campbell BC, Duffus JE, Osborne LS, Moran NA (1992) The eubacterial endosymbionts of whiteflies (Homoptera: Aleyrodoidea) constitute a lineage distinct from the endosymbionts of aphids and mealybugs. Curr Microbiol 25:119–123
5. Clark MA, Baumann L, Thao ML, Moran NA, Baumann P (2001) Degenerative minimalism in the genome of a psyllid endosymbiont. J Bacteriol 183:1853–1861
6. Costa HS, Westcot DM, Ullman DE, Rosell R, Brown JK, Johnson MW (1995) Morphological variation in *Bemisia* endosymbionts. Protoplasma 189:194–202
7. Douglas AE (1998) Nutritional interactions in insect-microbial symbioses. Annu Rev Entomol 43:17–37
8. Gullan PJ, Martin JH (2003) Sternorrhyncha (jumping plant-lice, whiteflies, aphids and scale insects). In: Resh VH, Cardae RT (eds) Encyclopedia of insects. New York: Academic Press/Elsevier Science, pp 1079–1089
9. Moran NA, Baumann P. (2000) Bacterial endosymbionts in animals. Curr Opin Microbiol 3:270–275
10. Munson MA, Baumann P (1993) Molecular cloning and nucleotide sequence of a putative *trpDC(F)BA,* operon in *Buchnera aphidicola* (endosymbiont of the aphid *Schizaphis graminum*). J Bacteriol 175:6426–6432
11. Pang KM, Knecht DA (1997) Partial inverse PCR: a technique for cloning flanking sequences. Bio Techniques 22:1046–1048

12. Shigenobu S, Watanabe H, Hattori M, Sakaki Y, Ishikawa H (2000) Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera* sp. APS. Nature 407:81–86
13. Spaulding AW, von Dohlen CD (2001) Psyllid endosymbionts exhibit patterns of co-speciation with hosts and destabilizing substitutions in ribosomal RNA. Insect Mol Biol 10:57–67
14. Stover CK, Pham XQ, Erwin AL, Mizoguchi SD, Warrener P, Hickey MJ, Brinkman FS, Hufnagle WO, Kowalik DJ, Lagrou M, Garber RL, Goltry L, Tolentino E, Westbrock-Wadman S, Yuan Y, Brody LL, Coulter SN, Folger KR, Kas A, Larbig K, Lim R, Smith K, Spencer D, Wong GK, Wu Z, Paulsen IT, Reizer J, Saier MH, Hancock RE, Lory S, Olson MV (2000) Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen. Nature 406:959–964
15. Tamas I, Klasson L, Canbäck B, Näslund AK, Eriksson A-S, Canbäck B, Wernegreen JJ, Sandström JP, Moran NA, Andersson SGE (2002) 50 million years of genomic stasis in endosymbiotic bacteria. Science 296:2376–2379
16. Thao ML, Moran MA, Abbot P, Brennan EB, Burckhardt DH, Baumann P (2000) Cospeciation of psyllids and their primary prokaryotic endosymbionts. Appl Environ Microbiol 66:2898–2905
17. Thao ML, Gullan PJ, Baumann P (2002) Secondary (γ-*Proteobacteria*) endosymbionts infect primary (β-*Proteobacteria*) endosymbionts of mealybugs multiple times and coevolve with their hosts. Appl Environ Microbiol 68:3190–3197
18. Thao ML, Baumann L, Hess JM, Falk BW, Ng JCK, Gullan PJ, Baumann P (2003) Phylogenetic evidence for two new insect-associated chlamydia of the family *Simkaniaceae*. Curr Microbiol 47:46–50
19. van Ham RCHJ, Kamerbeek J, Palacios C, Rausell C, Abascal F, Bastolla U, Fernandez JM, Jimenez L, Postigo M, Silva FJ, Tamames J, Viguera E, Latorre A, Valencia A, Moran F, Moya A (2003) Reductive genome evolution in *Buchnera aphidicola*. Proc Natl Acad Sci USA 100:581–586
20. Wernegreen JJ (2002) Genome evolution in bacterial endosymbionts of insects. Nat Rev Genet 3:850–861
21. Zchori-Fein E, Brown JK (2002) Diversity of prokaryotes associated with *Bemisia tabaci* (Gennadius) (Hemiptera: Aleyrodidae). Ann Entomol Soc Am 95:711–718
22. Zchori-Fein E, Gottlieb Y, Kelly SE, Brown JK, Wilson JM, Karr TL, Hunter MS (2001) A newly discovered bacterium associated with parthenogenesis and a change in host selection behaviour in parasitoid wasps. Proc Natl Acad Sci USA 98:12555–12560